The Hubble Space Telescope Engineering Data Warehouse

Presented By: Ken Lehtonen
NASA/Goddard

Agenda

- Introduction: "Quick Tour" of Hubble's Universe
- Overview of HST Reengineering Project
- Application of Data Warehouse Technology
- Progress To Date
- Remaining Challenges
- Lessons Learned
- For More Information

Comparison before and after HST Servicing Mission 1

M100 Galactic Nucleus

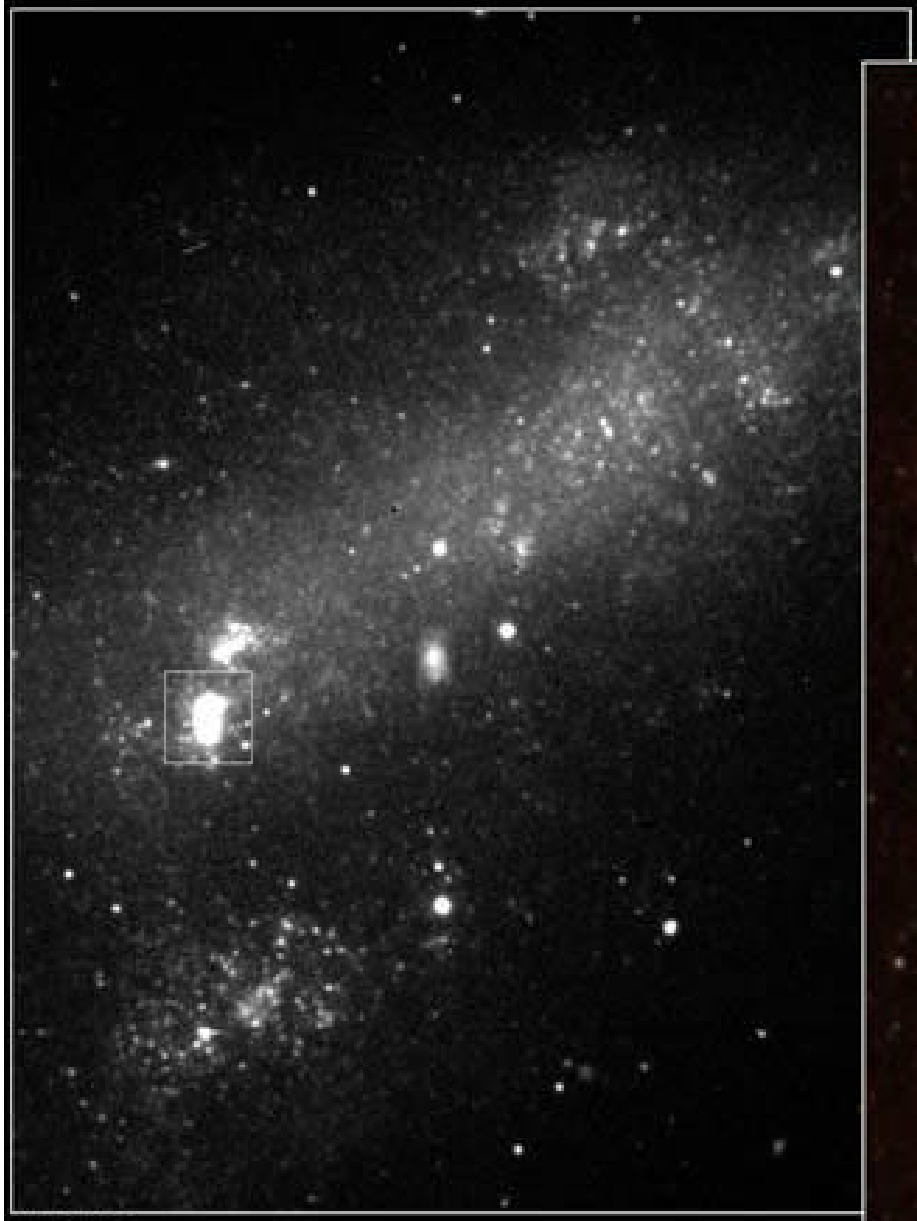
Wide Field Planetary Camera 1

Wide Field Planetary Camera 2

STScI-PR94-01


You are here
We are here

Starbirth Region in M33 Nebula, NGC604

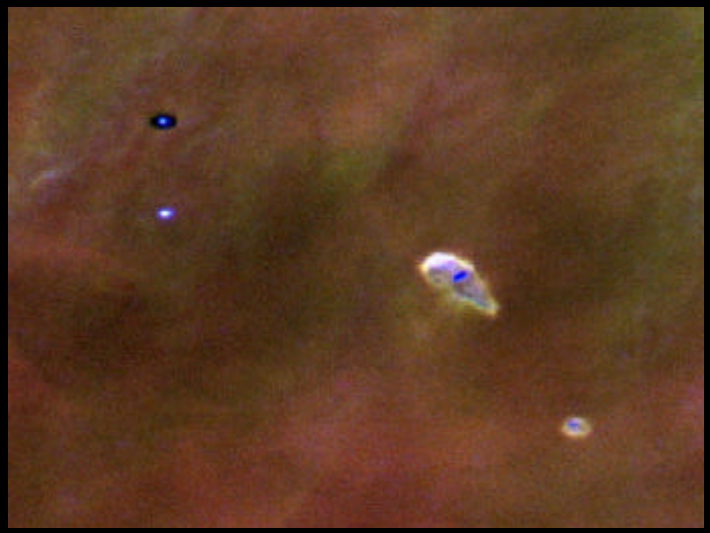


Mt. Polomar Observatory

STScI-PRC96-27

The Orion Nebula

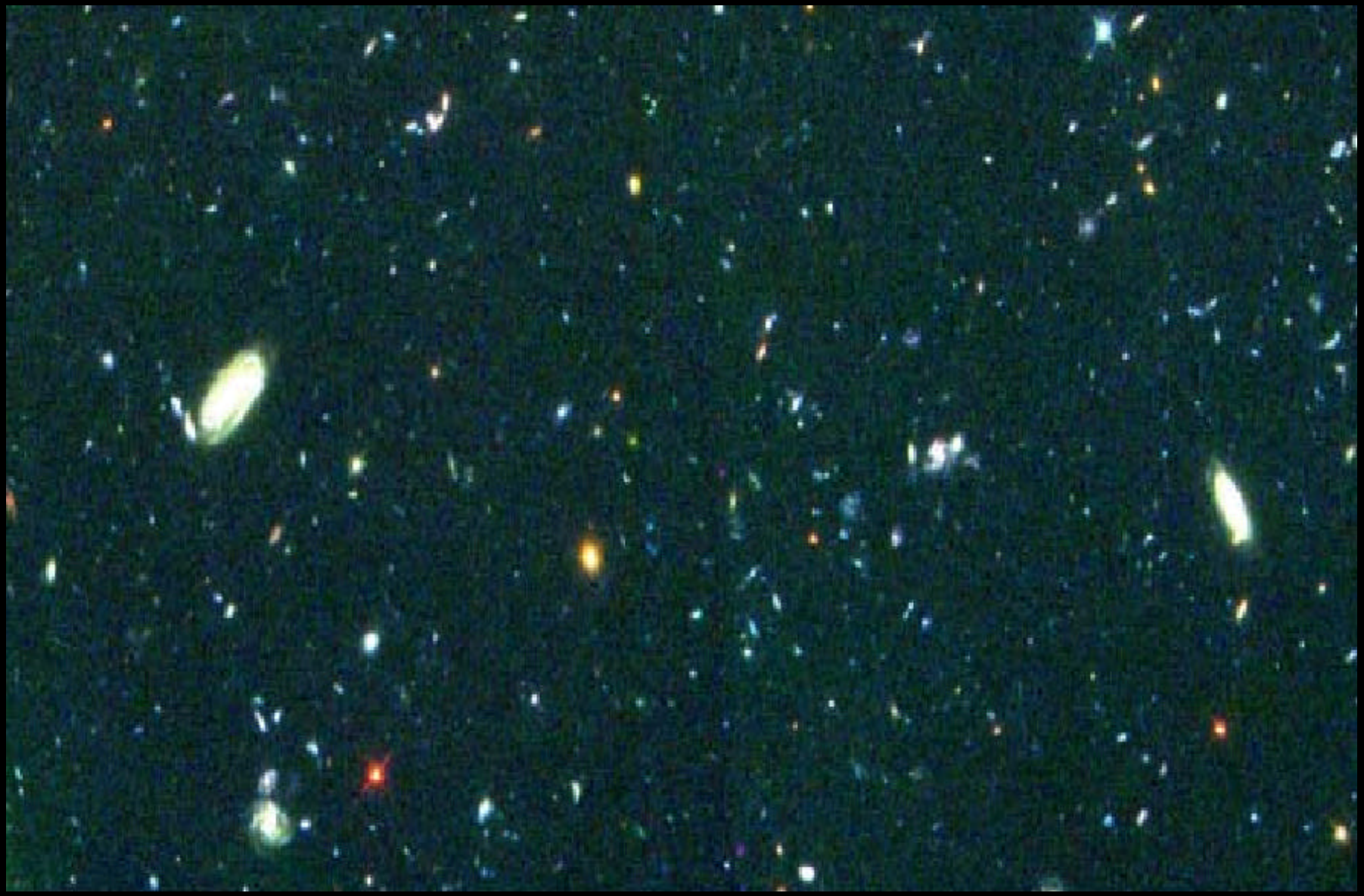
STScI-PRC95-45a



Doug Johnstone - CITA

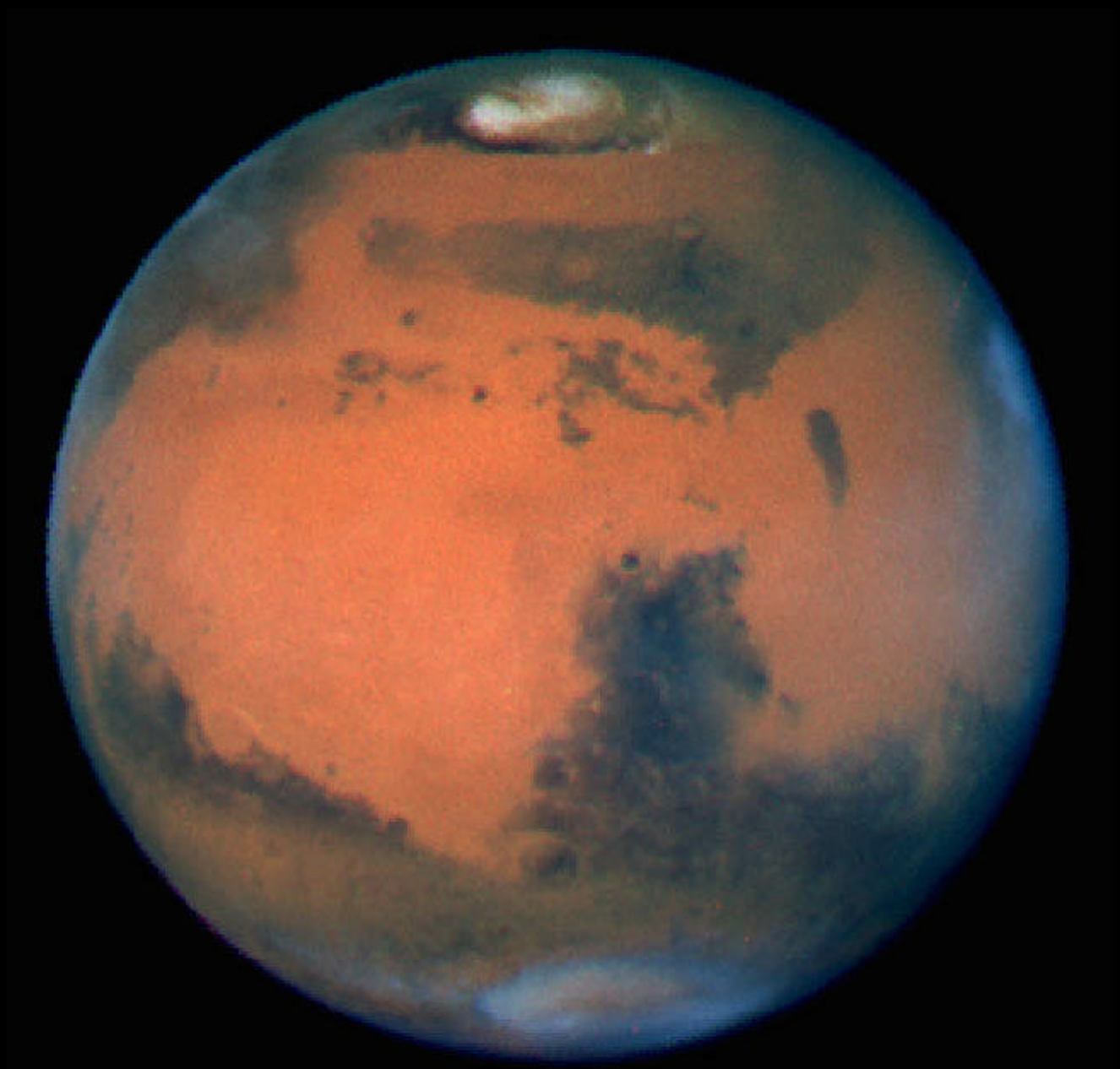
STScI-PRC95-45b

HST Young Galaxy Survey



STScI-PRC96-29a

Mars



STScI-PRC97-09a

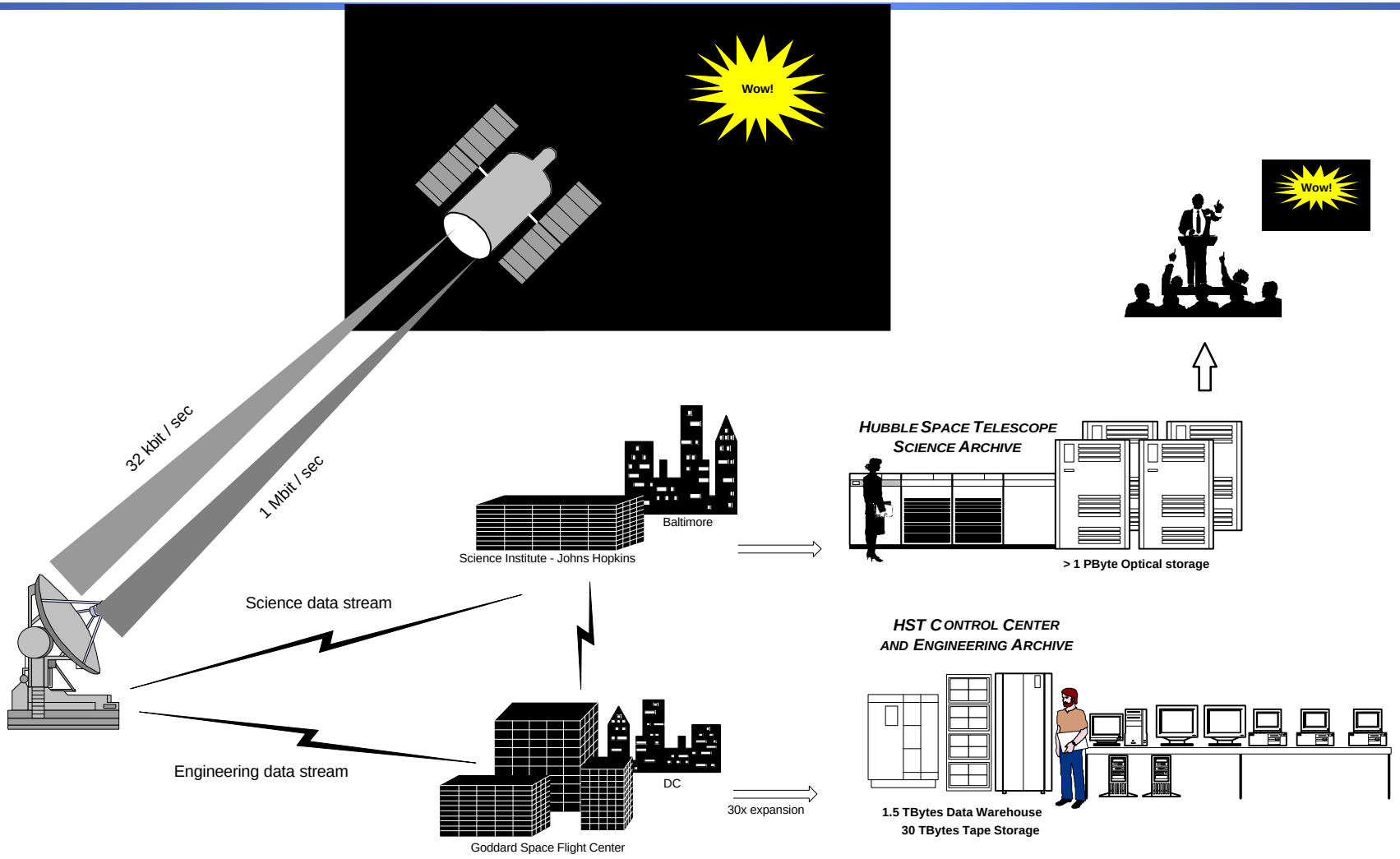
Overview of HST Reengineering Project "Mission Statement"

- Re-engineer the HST operational systems to streamline and modernize HST's overall ground system operations
- Reduce overall operations and maintenance cost of the system by at least 50% through automation of routine functions and use of off-the-shelf components
- Provide for monitoring the ground and space systems by the system itself, involving operations personnel *on an exception basis only*
- Provide an integrated system that replaces the functionality of currently disparate systems and data "marts"
- Support access to engineering telemetry data necessary to perform diagnosis from anywhere in the world
- Be a technological leader in ground systems by incorporating the latest proven technology and use, to the maximum extent practical, Commercial Off-The-Shelf (COTS) products (e.g., data warehouse)

Overview of HST Reengineering Project "The Control Center System (CCS)"

- Commands and controls the Hubble Space Telescope
- Processes and stores all of the *engineering* telemetry data:
  - spacecraft temperature
  - electrical power
  - attitude
  - on-board computer dumps, etc.
- Tracks all "vital signs" of the HST in a mission-critical sense
- Is evolving toward "lights-out"operations [or at least "dimmed"]

Elements of the End-to-End System



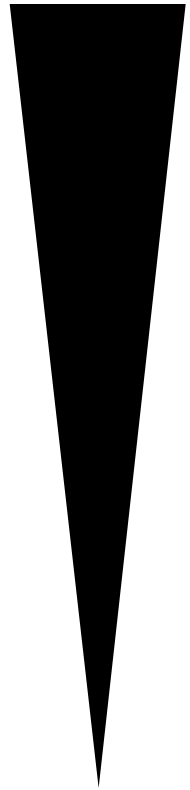
Application of Data Warehouse Technology "Goals & Objectives"

- Centralize all of HST's multiple data stores into one data warehouse
- Provide immediate access to all of HST's engineering telemetry data
- Enable user to issue "canned," customizable queries or to pose complex questions of the data using standard Structured Query Language statements or commercial tools
- Reduce the amount of time to analyze/predict anomalies with the HST spacecraft subsystems
- Provide access to engineering telemetry for a spacecraft with:
  - Full-time, 24x7 operations
  - A mission life expectancy to year 2010 (with two more servicing missions scheduled)

Application of Data Warehouse Technology "Early Work"

- Conducted product searches; downloaded demo software; and, contacted prospective vendors
- Evaluated several data warehouse products and conducted performance benchmarks on each: *benchmarking was critical*
- Selected **Red Brick Systems** and initiated a small pilot phase with the Space Telescope Science Institute
  - Retained the services of a Red Brick consultant to assist in the design of the schema and to work issues with Corporate
  - Held several vendor training classes for the team including DBAs
- Incorporated results of pilot phase into the overall architecture and design of the CCS for a future delivery. Pilot phase looked at DW structure and compared:
  - segmentation by time (months/years) vs.
  - segmentation by mnemonic

The Analysis Spectrum



General	Data Warehouse	Discrete State
	Abstract Analysis	***Scientist***
	relational operations	
	count occurrences	***Instrument Engineer***
	correlate events	
	Summary Analysis	
	long-term trends	***Operations***
	averages	
	Detailed Analysis	***Spacecraft Engineer***
	short periods of time	
	high granularity	
Specific	***All-points flat files***	***Time Continuum***

Application of Data Warehouse Technology "Solving the Data Problem"

- Problem In A Nutshell: *"In one year, HST engineering stream produces 1.5 TB in flat files or 100 billion rows in a database."*
- Solution: "*Implement Data Reduction techniques*":
  - Algorithmic Data Reduction: using "change-only" and statistical sampling, 100 B rows/year reduced to 1.8 B rows/year + indices and stored in the DW on high-density RAID disk drives
  - Data Compression: 1.5 TB/year compressed to 500 GB/year and stored in the flat-file archive using DLT jukebox technology
    - *"DATA FROM LAUNCH THROUGH YESTERDAY IS ONLINE"*
- Once the data has been cleansed and "pre-loaded", the Red Brick Product will:
  - Load 24 hours of data in 15 minutes!!
  - Segment the data by month
  - Store the data into files based on time
  - Support fast queries via "star indices"

Application of Data Warehouse Technology "A New Operations Concept"

- Data Warehouse manages aggregate data
  - is used for majority of analysis by spacecraft engineers:
    - » trends, signatures, relations, or data mining
    - » time-based analysis at low resolution, on the order of days to weeks
    - » tool to focus analysis from general to specific
- Data Warehouse becomes index into flat files
  - flat-file requests should be very specific:
    - » time resolution from milliseconds to hours
    - » few to several data items
- **Summary:**
  - Constructed the data warehouse for 80% of our analyses
  - Preserved 100% of our data in an archive
  - Used a common index: time (GUI or data server bridges the data sets)

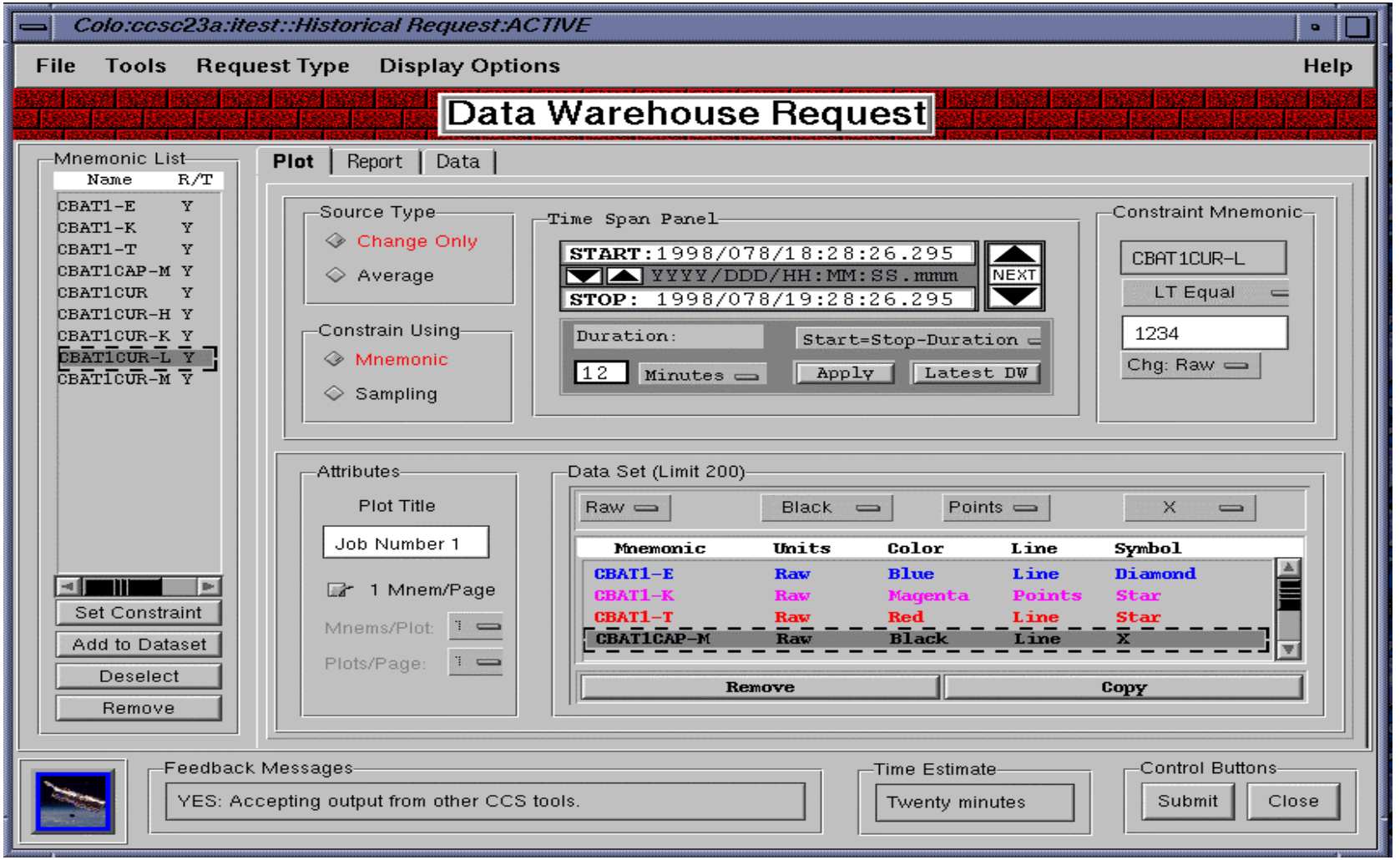
Application of Data Warehouse Technology "Current Schema Contents"

- Engineering Telemetry
  - **Averaged, high rate**: Mnemonic, Start/Stop Time, Min, Max, Avg, Telemetry Format, Sample Count, Status flags
  - **Change Only, low rate**: Mnemonic, Start/Stop Time, Telemetry Format, Discrete Value, EU value, Raw Value, Status flags
- Events
  - Spacecraft and Orbital Events, planned, not actual
  - Used to find/describe telemetry
- Metadata
  - Mnemonic definitions and descriptive data ("definitions & attributes")
  - Discrete codes for "Discrete & Binary" mnemonics ("valid values")
  - Event Class descriptions and defaults
- Dimensions
  - Date, Time, Millisecond

Progress To Date

- Schema Finalized for Telemetry
  - Simple
  - Narrow Rows: indices pointing to data
- Finalizing Design For Other Data (separate tables)
  - Events (day/night crossings; South Atlantic Anomaly)
  - Shuttle/Servicing Mission Data
  - Derived Mnemonics
- Installed the Red Brick Product
- Established Data Warehouse Working Group
  - Spacecraft Engineers
  - Representatives from operations
- Executed, via the GUI, Several Canned, Customizable Queries Against the Telemetry To Establish Performance Baseline

Application of Data Warehouse Technology
"Sample GUI Data Request Screen"



Remaining Challenges

- Integrate the Data Warehouse product into the CCS development cycle and perform initial checkout of integrated system: complete by April 30
- Verify query capabilities (simple & complex) and perform associated performance testing with live users and real data
- Deliver the system for operational support commencing August 31
  - Current data merged into 24 hour data sets then filtered and reduced using a custom preloader utility (in 2 1/2 hours)
  - Data Warehouse utility loads 5 million rows per day (in 15 minutes)
- Begin loading eight year's worth of legacy telemetry data into warehouse commencing September 15 and completing one year later
  - Use separate system to process the legacy data (synchronize with current process)
  - Plan to load 1 TB of data in one year
- Become experienced with the data warehouse in an operational setting

Lessons Learned

- Implementation requires that upfront design and requirements work be much more complete than a conventional data management project
- Benchmarks of potential vendor products should be performed
- A pilot project should be undertaken ("build in small steps") and then reviewed by potential users
- JAD sessions are useful in "setting expectations" as well as capturing real user requirements (evolutionary process)
- Consultants (from the selected vendor) can be very helpful during the design and implementation process:
  - accelerate the design process itself (schema development)
  - are invaluable in understanding the actual data warehouse product

Lessons Learned (continued)

- The more open the architecture, the easier it is to incorporate new tools (e.g., reporting tools) as they become available
- Important to train your own staff ("more cost effective") as experienced DW designers and DBAs are in short supply
- Data Warehouses are "pricey": need unqualified support from upper management and associated commitment in terms of resources: $$ and people (Note: ROI often in terms of *value*.)

For More Information

- E-Mail: **klehtonen@v2pop.hst.nasa.gov**
  Voice: (301) 918-7412

- Results of the Data Warehouse Pilot Study
  http://www.sesd.stsci.edu/ccs_dw/

- CCS URL: **http://ccs.hst.nasa.gov**
  - Point to “Data Management” Team
  - Scroll for “CCS Data Warehouse Pilot Study”

- Red Brick’s URL: **http://www.redbrick.com**

- DM Review’s URL: **http://www.data-warehouse.com/**